

试论 C 语言的字符数据

赵 静, 宋灵香

(定西师范高等专科学校 计算机系, 甘肃 定西 743000)

【摘 要】在编写 C 程序时经常会用到字符数据和字符串, 很多人在进行程序设计时很容易混淆字符和字符串的处理方法, 因此笔者从字符和字符串的定义、存储、输入和输出方法几个方面阐述其异同, 在程序设计时能正确使用字符或字符串解决实际问题。

【关键词】字符; 字符串; 存储

【中图分类号】TP312 **【文献标识码】**A **【文章编号】**1673-1891(2014)02-0074-03

引言

在编写 C 程序时经常会用到字符数据。C 语言的字符数据总结起来只有两种: 字符和字符串, 但很多人对它的认识总是不够清晰, 导致写程序时处于模棱两可的状态, 常常会产生不少错误。只有对字符数据有一种清晰而明确的认识, 才有可能写出正确、安全和高效的 C 程序。本文根据字符和字符串的基本内容及区别进行总结, 并结合一些典型例子, 对字符数据做一阐述。

1 字符和字符串的定义

C 语言中的字符型数据, 分为字符常量和字符变量, 字符常量是用单引号(即撇号)引起来的一个字符, 代表 ASCII 字符集中的一个字符。例如: 'a', 'A' 等。字符变量是用来存放字符常量的, 只能放一个字符。而在 C 语言中并没有字符串数据类型, 却允许使用“字符串常量”, 但没有“字符串变量”, C 语言中的字符串常量, 是用双引号括起来的零个或多个字符序列, 因此像 " ", "a", "How are you!", "1234abcd" 等都是字符串, 串中可以包含普通字符, 也可以包含转义字符(由 \ 加特定字符组成的具有特殊意义的字符常量)。字符变量不能存放字符串常量。

2 字符和字符串的区别

要明确区分字符和字符串就要清楚它们之间的异同, 从以下几方面加以说明:

2.1 定义方法

字符是由单引号引起来的一个字符, 而字符串是由双引号引起来的一串字符。字符只能是一个字符, 字符串可以是零个或多个字符。例如: 'a' 和 "a" 虽然都只有一个字符, 但前者是字符常量, 而后者是字符串常量。

2.2 字符和字符串的存储方式^[4]

字符的存储非常简单, 它只有一个字符, 在内存中占一个字节, 将字符常量存放到字符变量中, 并不是将该字符本身放到内存单元中, 而是将该字符的相应的 ASCII 代码放到存储单元中。例如字符 'a' 的 ASCII 代码为 97, 在内存中相应变量 C1 的值, 实际是以二进制形式存储的, 如图 1 所示。

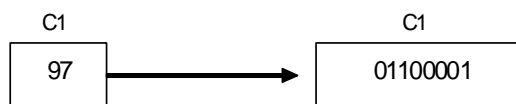


图 1

字符串的存储则比字符要复杂的多, 在 C 语言中没有简单变量可以存放字符串, 只能用字符型数组存放字符串, 字符串在内存中存储时, 除了存储字符串的有效字符之外, C 规定在字符串结尾加 '\0' 作为字符串的结束标志, 以便系统据此判断字符串是否结束。因此遇到 '\0' 认为字符串结束(串的长度中不包含 \0), 而不在去考虑 '\0' 后的字符, 这说明双引号中的内容不一定是字符串中的内容。例如 "CHINA", 在内存中的存放是这样的, 如图 2 所示。



图 2

注: 在写字符串时不必加 '\0', 否则会画蛇添足, '\0' 是系统自动加上的。

2.3 字符和字符串的结束标志

字符没有结束标志, 因为它只有一个字符, 而字符串在内存中存放时系统自动加 '\0' 作为结束符。'\0' 是一个 ASCII 代码为 0 的字符, 从 ASCII 代码表中可以看到 ASCII 代码为 0 的字符是“空操作字

收稿日期: 2014-03-25

作者简介: 赵静(1978-), 女, 甘肃省定西市人, 本科, 讲师, 研究方向: 计算机应用。

符”,它不引起任何控制动作,也不是一个可显示的字符,不做任何操作,只是一个标志。它不计入串的长度。因此,在字符数组中如存有‘\0’,则认为是存放了字符串,可按字符串进行操作;如没存放‘\0’,则不认为是存放了字符串,例如:

```
Char a [5] = { 'a', 'b', 'c', '\0' };
```

```
Char b [3] = { 'a', 'b', 'c' };
```

则数组a中存放了字符串"abc",而数组b中存放的不是字符串"abc",而是三个字符‘a’,‘b’,‘c’。

2.4 字符和字符串的长度^[2]

由定义可知,字符只能是一个字符,因此它的长度肯定是1,而字符串的长度是指字符串中所包含字符的个数,字符串结束符‘\0’不计入字符串的长度。像""、“a”、“abc23”等串的长度计算非常容易,它们的长度分别为0、1、5,但对于一些像"v66.45\rdel023\N\088av"、“asdf\09812a”这样的串,往往让很多人无从下手计算长度。在C语言常见的字符串中,可以归纳为以下几种情况:

2.4.1 由一些常用的普通字符组成的字符串

如"as123",它的存储形式为:

a	s	1	2	3	\0
---	---	---	---	---	----

串的长度为5,即字符串中包含的所用字符的个数。不包含“\0”。

2.4.2 字符串中包含转义字符,一个转义字符算一个字符

如:" ab1cfn\b",它的存储形式为:

a	b	1	c	f	\n	\b	\0
---	---	---	---	---	----	----	----

串的长度为7。常见的转义字符有\n(换行)、\f(换页)、\r(换行符)、\t(水平制表符)、\b(退格)、\v(垂直制表符)、\'(单引号)、\"(双引号)、\\(斜杠字符“\”)、\ddd(1到3位八进制所代表的ASCII码)、\xhh(1到2位十六进制所代表的ASCII码)等。如果在字符串中出现这些字符时,每出现一次算一个字符。

2.4.3 字符串中包含类似转义字符的序列字符

字符串中包含类似转义字符的字符时,一定要能明确的区分些字符是否是转义字符,否则就会产生错误。如:串" g1.23\023\rbc\Nu\088ef",它在内存中的存储形式为:

g	1	.	2	3	\r	b	c	N	u
---	---	---	---	---	----	---	---	---	---

其中有\n、\088时。\\N不构成转义字符,输出串时,\将被忽略,只计N一个字符;\088不是任何

一个字符的八进制或十六进制ASCII代码值,系统认为遇到‘\0’则字符串结束。因此串的长度为10。

3 字符和字符串的输入与输出^[3]

了解字符和字符串之后,在编写C程序时,往往会涉及它们的输入与输出方法。

3.1 字符数据的输入与输出

3.1.1 getchar 函数和 putchar 函数

getchar 函数;getchar 函数的作用是从终端(通常是键盘)输入一个字符,其函数值就是从键盘输入的字符,可将函数值赋给字符变量,注意 getchar 函数只能接收一个字符,没有参数。如:

```
char c1;
```

```
c1=getchar();
```

putchar 函数;putchar 函数的作用是向终端输出一个字符,这个字符可以是字符变量,也可以是字符常量或转义字符。如:putchar(c1)、putchar(‘a’)、putchar(‘\n’)、putchar(‘\101’)等。

3.1.2 gets 函数和 puts 函数

gets 函数;gets 函数的作用是从终端输入一个字符串到字符数组中。如:char str1[13];

```
gets(str);
```

puts 函数;puts 函数的作用是将一个字符串(以‘\0’结束的字符序列)输出到终端,在输出时将字符串结束标志‘\0’转换成‘\n’,即输出字符串后换行。

3.1.3 scanf 函数和 printf 函数

scanf 函数;scanf 函数是一种格式控制输入函数,可以用于输入任何类型的数据,也包含字符型数据,用scanf函数给字符变量赋值使用格式说明符“%c”,表示要输入一个字符。用回车表示输入结束,但用“%c”格式输入字符时,空格和“转义字符”都会作为有效字符输入,因此使用时要特别注意。用scanf函数为字符数组赋值时,需用“%s”做格式说明符。如:

```
char str1[10],str2[10],str3[10];
```

```
scanf(“%s%s%s”,str1,str2,str3);
```

printf 函数;printf 函数是格式输出函数,输出单个字符时,使用格式说明符“%c”,如:

```
char c1=‘A’,c2=‘B’;
```

```
printf(“%c,%c”,c1,c2);
```

若要输出字符数组,可以用格式符“%c”或“%s”。用“%c”输出字符数组,必须利用循环来完成输出。若使用格式符“%s”,意为输出字符串,在输出

时遇到字符串结束标志‘\0’就会停止输出。在用 printf 函数的“%s”格式输出字符串时不会将字符串结束标志‘\0’转换成‘\n’,如果需要换行必须加一条换行语句 printf (“\n”);

3.2 输入和输出举例

3.2.1 用 scanf()输入字符串,printf()输出字符串:

逐个字符处理: #include "stdio.h "

```
void main()
{ char ch[10]; int i;
for(i=0;i<9;i++)
scanf("%c",&ch[i]);
for(i=0;i<9;i++)
printf("%c",ch[i]);}
```

整个字符串处理: #include "stdio.h "

```
void main()
{ char ch[10];
scanf("%s",ch);
printf("%s",ch);}
```

说明:以字符串为单位处理时,直接使用数组名,无需&;输入字符串时,字符个数要小于数组的长度,如输入5个字符,定义的字符数组至少应有6个元素;输入字符串时,若遇空格或回车,输入结束,并自动在串后加上结束标志‘\0’;输出字符串时,遇到字符串结束标志‘\0’,输出结束。

3.2.2 用字符串处理函数输入和输出

注释及参考文献:

[1]张岩.C语言中的字符串[J].牡丹江师范学院学报,2003:6-8.
 [2]白春霞,张维宝.基于C语言中字符串长度计算问题的探讨[J].电脑知识与技术,2010.(10):8216-8217.
 [3]崔琦.浅议C语言中字符数据的输入与输出[J].中国科技信息,2009(14):117-118.
 [4]谭浩强.C程序设计(第三版)[M].北京:清华大学出版社.2005.

字符串标准函数的原型在头文件 string.h 中(使用字符串处理函数输入和输出时要在头文件上加入 #include<string.h>) 字符串输出函数 puts ,字符串输入函数 gets

例如:

```
#include<stdio.h> #include<string.h>
void main() {
char str[20]; int i;
printf("请输入字符串:\n"); gets(str);
printf("输入的字符串为:\n"); puts(str);
}
```

3.3 C 程序设计中有字符串的一些常见错误

- (1)字符数组定义时没有为字符串结束符‘\0’开辟存储空间;
- (2)输出了不包含字符串结束符‘\0’的“字符串”;
- (3)混淆处理字符串和字符的方法;
- (4)在函数的参数传递时字符串和字符不统一,即字符串参数传递了字符或字符参数传递字符串。

4 结束语

本文只是对C语言的字符数据做了一个简单的概括,如要对其熟练的掌握,除了需要在概念上有一个明确清晰的认识外,还需要在实际的编程中多加练习摸索总结,达到熟练的应用。

Character Data in C Language

ZHAO Jing, SONG Ling-xiang

(Department of Computer, Dingxi Teachers College, Dingxi, Gansu 743000)

Abstract: Frequently used character data and string when writing a c program, a lot of people easily confuse the character and string handling methods at the time of programming.This paper elaborates the similarities and differences among the characters and the string definition, methods of storage, input and output so as to solve practical problems through the character or string in program design.

Key words: character; string; storage